Differential coexpression analysis of obesity-associated networks in human subcutaneous adipose tissue

Published in final edited form as:
Int J Obes (Lond). 2012 January ; 36(1): 137–147. doi:10.1038/ijo.2011.22.

Differential co-expression analysis of obesity-associated networks in human subcutaneous adipose tissue

A.J. Walley[1,*], **P. Jacobson**[2,*], **M. Falchi**[1,*], **L. Bottolo**[1,3], **J.C. Andersson**[1,2], **E. Petretto**[3,4], **A. Bonnefond**[5], **E. Vaillant**[5], **C. Lecoeur**[5], **V. Vatin**[5], **M. Jernas**[2], **D. Balding**[1,6], **M. Petteni**[1], **Y.S. Park**[1], **T. Aitman**[4], **S. Richardson**[3], **L. Sjostrom**[2], **L. M. S. Carlsson**[2,*], and **P. Froguel**[1,5,*,†]

[1] Department of Genomics of Common Disease, School of Public Health, Imperial College London, Hammersmith Hospital, Du Cane Road, London, W12 0NN, UK

[2] Department of Molecular and Clinical Medicine, The Sahlgrenska Academy, Gothenburg University, SE-413 07 Gothenburg, Sweden

[3] Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, St Marys Hospital, 161 Norfolk Place, London, UK

[4] MRC Clinical Sciences Centre, Division of Clinical Sciences, Imperial College London, Commonwealth Building, Hammersmith Hospital, Du Cane Road, London, W12 0NN, UK

[5]CNRS 8090-Institute of Biology, Pasteur Institute, Lille, France.

[6]Institute of Genetics, University College London, Kathleen Lonsdale Building, 5 Gower Place, London, WC1 E6B, UK

Abstract

Objective—To use a unique obesity-discordant sib-pair study design to combine differential expression analysis, expression quantitative trait loci (eQTLs) mapping, and a co-expression regulatory network approach in subcutaneous human adipose tissue to identify genes relevant to the obese state.

Study design—Genome-wide transcript expression in subcutaneous human adipose tissue was measured using Affymetrix U133+2.0 microarrays and genomewide genotyping data was obtained using an Applied Biosystems SNPlex linkage panel.

Subjects—154 Swedish families ascertained through an obese proband (Body Mass Index >30kg/m^2) with a discordant sibling (BMI>10kg/m^2 less than proband).

Results—Approximately one-third of the transcripts were differentially expressed between lean and obese siblings. The cellular adhesion molecules (CAMs) KEGG grouping contained the largest number of differentially expressed genes under *cis*-acting genetic control. By using a novel approach to contrast CAMs co-expression networks between lean and obese siblings, a subset of differentially regulated genes was identified, with the previously GWAS obesity-associated *NEGR1* as a central hub. Independent analysis using mouse data demonstrated that this finding for *NEGR1* is conserved across species.

[†] Corresponding author. Department of Genomics of Common Disease, School of Public Health, Imperial College London, Hammersmith Hospital, Du Cane Road, London, W12 0NN, UK p.froguel@imperial.ac.uk., tel +44 (0)20 759 46520.
[*] these authors contributed equally to this work

Conflict of Interest
The authors declare no competing financial interests.

Europe PMC Funders Author Manuscripts

Conclusion—Our data suggests that, in addition to its reported role in the brain, *NEGR1* is also expressed in subcutaneous adipose tissue and acts as a central "hub" in an obesity-related transcript network.

Keywords

Gene Expression; network; eQTL; sibpair; linkage; adipose tissue

Europe PMC Funders Author Manuscripts

Introduction

Obesity, commonly defined as a body mass index (BMI) > 30 kg/m^2, has steadily risen in prevalence globally, a trend that could lead to over a billion people being obese by 2030 [1]. Obesity is already a major public health problem, resulting in increased morbidity and mortality [2] and different hypotheses have been suggested to account for this [3]. Genome-wide linkage analysis alone has identified many genomic regions linked to obesity but replication has been problematic [4]. More recently, common low-penetrant variants associated with obesity have been identified in genome-wide association studies (GWAS) [5-8]. Additionally, rare copy number variants [9] have also been implicated in the causality of obesity. All of these approaches rely on the correlation between genomic variation and either obesity status or an obesity-related quantitative phenotype, e.g. BMI.

Gene expression levels reflect the combined effects of a wide range of genomic modifications including point mutations, structural variants and epigenetic changes. Abundance of any specific mRNA is therefore likely to more closely reflect the overall genomic effects than each type of variation separately. This is especially true for those changes having a direct effect on the transcription levels, although alterations in protein structure and function might also have a feedback effect on transcriptional activity [10]. Environmental effects are also likely to be indirectly captured by transcript levels, as recently shown in leukocyte gene expression studies among three Moroccan sub-populations where at least 37% of the differentially expressed transcripts were not explainable by genetic and methylation differences[11]. Therefore, the assessment of genome-wide gene expression provides a snapshot of underlying cellular processes and their environmental and genomic influences.

Since the transcript levels are strongly modulated by polymorphisms in regulatory regions, they can be powerfully mapped by correlating gene expression with genetic data. The regions identified by such correlations, named expression quantitative trait loci (eQTLs), directly pinpoint the functional link between variants in the genome and their biological effect. For this reason, eQTL analysis has been suggested as a means to identify genetic variants involved in the susceptibility to complex diseases and to fill the gap between disease associations identified by GWA and the mechanism by which they contribute to the disease [12, 13]. The choice of tissue is central to a gene expression study, as the expression profile is context dependent and differs between tissues [14]. In addition, within the same tissue, eQTLs can be specific to the cellular differentiation state [15]. Subcutaneous adipose tissue (SAT) is the tissue of choice to investigate common human obesity because it displays obesity-related changes in gene expression [16], it has clear endocrine organ characteristics [17], and samples can be obtained from large numbers of human subjects. Altered expression of a number of genes implicated with obesity and the metabolic syndrome has been reported in studies of SAT from obese subjects, including *CD36* [18] and *PFKFB3* [19].

Instead of analyzing each transcript independently from the others, novel approaches can exploit the interactions among transcripts to identify gene networks. They delineate the complex interrelationships occurring amongst gene transcription levels which can be

correlated with phenotypic and genomic data for the identification of relevant biological pathways [12]. Measurement of gene expression in multiple tissues in mice has allowed the delineation of a gene network enriched for genes involved in the inflammatory response and macrophage activation that is highly correlated with obesity-related phenotypes [20]. A similar overlapping network has been identified in human SAT [21].

Our study takes advantage of the SibPair cohort, which consists of154 families (n=732) identified by having an obese proband (BMI>30kg/m^2) with a BMI-discordant sibling (BMI difference of at least 10kg/m^2) [22]. SAT and blood samples were available from the siblings and peripheral blood from all subjects. These unique discordant families allowed a combined approach for the identification of genes and pathways involved in obesity. Using a relatively small sample, we have combined eQTL mapping, differential-expression analysis, and a novel differential co-expression network approach in sib-pairs to identify biologically-relevant transcriptional modules and their key regulators to provide insights into the pathogenesis of obesity.

Materials and Methods

Participants and study design

The study cohort was 154 nuclear families (732 subjects) ascertained via an extremely BMI-discordant sib-pair (difference >= 10kg/m^2)[22]. Average family size was 4.75. SAT samples were available from the siblings and peripheral blood from all subjects. Median BMI (1st-3rd quartiles) was 27.2 (23.0–33.2), range 16.9-57.8. Median age (1st-3rd quartiles) was 45 years (36–63). Informed written consent was obtained from all participants. This study was approved by the ethics committee of Gothenburg University.

Nucleic acid isolation

Genomic DNA was isolated from whole blood using the QIAamp DNA Blood Maxi Kit (Qiagen, Hilden, Germany) according to the manufacturers' recommendations. Subcutaneous adipose tissue biopsies were immediately frozen in liquid nitrogen and RNA was extracted using the Qiagen RNeasy Lipid Tissue kit.

Linkage Genotyping

The SNPlex™ System Linkage Mapping Set (http://www.appliedbiosystems.com) was used, comprising 3 922 SNPs, of which ~75% are in clusters, distributed across 95 probe pools. Allelic discrimination was performed using an Applied Biosystems 3730xl DNA Analyzer and GeneMapper3.7 software. Pedcheck [23] was used to detect Mendelian inconsistency. Genetic markers giving rise to tight double recombinants were identified with MERLIN [24] and treated as missing data.

Gene expression measurement

Gene expression was measured using the Affymetrix Human Genome U133+2.0 array (Affymetrix, Santa Clara, CA). In brief, RNA was reverse transcribed into cDNA and biotin-labelled cRNA was prepared by *in vitro* transcription (Enzo Diagnostics Inc, Farmingdale, NY). After hybridisation, the arrays were scanned using the Affymetrix GeneArray GCS3000 scanner and visualised using GeneChip Operating Software (GCOS, Affymetrix). Gene expression levels were normalised using the Robust Multiarray Average (RMA) method [25].

RT-PCR gene expression analysis

Adipose tissue biopsies were obtained from subcutaneous fat depots of two French volunteers, as previously described [26]. For each sample, 1μg of total RNA was transcribed into cDNA using the cDNA Archive Kit (Applied Biosystems) or Random Primed First Strand Synthesis (Applied Biosystems). 4μl of a $1/10^{th}$ dilution of each resulting cDNA was used in a 20μl reaction, including 10μl of TaqMan gene expression mastermix (Applied Biosystems) and 1μl of the appropriate assay (Applied Biosystems). Quantitative RT-PCR analyses were performed using ABI 7900 HT SDS2.3 software and each sample was run in triplicate. *NEGR1* expression levels were obtained relative to three housekeeping genes (*ACTB*, *TOP1* and *POLR2A*). The cDNA sample content was normalized by subtracting the number of copies of the mean of three housekeeping genes from the number of copies of the target gene (ΔCt = Ct of target gene – Ct of housekeeping genes). Expression was calculated using the formula 100×2-ΔCt.

Linkage analysis

After quality control, 149 families were considered suitable for analysis. We selected the subset of transcripts having a unique position or specificity > 70% in the genome (n=27 904 transcripts) using SCAMPA (http://web.bioinformatics.ic.ac.uk/scampa). Linkage was evaluated using MERLIN-REGRESS[24]. Although robust to misspecification, MERLIN-REGRESS requires the population trait's mean, variance, and heritability. Population parameters were estimated using the variance component model implemented in SOLAR [27]. Since the variance components analysis requires a normal distribution for the trait, we applied a Box-Cox transformation to each transcript level [28]. Gene expression values falling outside the mean ±3 SDs were excluded from the analysis. Age and sex were included as covariates in the SOLAR analyses.

To identify *cis*-eQTLs, a window of 2.5cM left and right of each transcript position was used. Given this map resolution there are 1 483 transcripts which have no marker within 5cM, therefore a subset of 26 421 transcripts was analysed. All 27 904 transcripts were included in the *trans*-eQTL analysis. Linkage disequilibrium among the SNPs was modelled by specifying in MERLIN-REGRESS to treat as a “super-locus” all SNPs for which the observed pairwise $r^2>0.1$ [29]. All *P* values were calculated from LOD-scores, then corrected for multiple testing by the FDR procedure [30].

Assessing the significance of trans-eQTLs

To determine the empirical significance of *trans*-eQTLs, the approach of Emilsson [21] was used. Linkage analysis of the 27 904 transcripts was repeated using ten genome-wide datasets simulated by gene dropping under the null hypothesis of no linkage. The top-hit *trans*-eQTL for every transcript was extracted from each of the ten genome-wide analyses, giving a distribution of 279 040 LOD-scores that was used to assess empirical *P* values for the *trans*-eQTLs observed in the original data.

For the detection of hotspots of *trans*-regulation, we are interested in the probability for different signals, each of them genome-wide significant, to randomly arise at the same location. Hidden underlying correlation structure between the IBD at a genetic location and the transcription levels might influence the occurrence of false coincident linkages. The 5% LOD-score observed in the simulated dataset was used as threshold for the genome-wide significance of each analysed transcript in our data. The number of coincident linkages was then recorded at each marker location. Applying the same procedure to the simulated dataset, we obtained the distribution of coincident linkages under the null hypothesis of no-linkage. We used this distribution to assess empirical *P* values for the size of the observed

coincident linkages. Finally, multiple test correction was assessed using the FDR procedure [30].

Differential expression

Log-transformed expression levels for the whole set of 54 675 transcripts were corrected for age and sex and 119 pairs of extreme sibs were selected. The Limma package was used to identify significant genes that were over- or under-expressed [31]. Linear and robust regressions were performed separately, before applying the Empirical Bayes shrinkage method, obtaining similar results. Paired design was taken into account and specified accordingly. Correction for multiple testing was performed using Storey's FDR procedure [32] on the *P* values of the shrunk test statistics.

Differential co-expression analysis

Diseases can often result from the dysregulation of a gene network [33]. Differential co-expression analysis [34 35] might help in identifying those genes within the network that lead to the disruption of the regulatory mechanisms.

We propose a novel approach of testing the difference between gene networks in two groups. Firstly, we built obese and lean relevance networks with correlation matrices calculated using Kendall's *tau* correlation [36] in order to robustify the analysis. Then we contrasted the two networks calculating the differences between the transcript-transcript correlation matrices. Significant difference were evaluated using permutation tests [37] with different resample schemes chosen according to the two samples dependencies. Empirical *P* values were computed as the proportion of the differences observed in the permuted data sets that were equal or greater than what was observed in the original data set. An FDR thresholding procedure [32] was applied to the empirical *P* values to highlight the most significant differences.

Our approach, although similar in spirit to other methods that look at differences in coexpression networks between different conditions/or case control groups (for a review see [38]), is new in many respects. Firstly, through a model-free permutation test, we test directly whether the observed correlations differences are significant so we are not considering differences in the graph's topology [39]. Secondly, simply changing the sampling scheme for the permutation test, we can accommodate different levels of dependence between the groups. Thirdly, we do not consider just strong (positive or negative) correlations or strong differences using *ad hoc* thresholding [40]. Selection of what is relevant is obtained by applying the FDR procedure. Finally, the network module is defined as the connected component after FDR calculation, avoiding the *ad hoc* metric distances required in cluster algorithms [40], [41]

Identification of obesity-related biological pathways

At 10% FDR level we selected those differentially expressed transcripts for which *cis*-eQTLs were also identified. Enrichment of KEGG pathways was assessed with DAVID. Using all differentially expressed transcripts belonging to identified KEGG pathways and the same sub-sample selected by the Limma analysis, we applied the differential co-expression analysis approach at 10% FDR level. To take into account the paired design, we randomly relabelled the data within each pair in each permuted data set.

We tested whether the number of connections observed for the analysed genes was larger than that expected under the null hypothesis of these genes being randomly connected [42]. We also contrasted 1 000 relevance networks between obese and lean subjects generated using randomly-selected transcripts. The maximum number of connections was recorded for

each simulation to evaluate the empirical significance for the most connected genes in the original dataset.

Validation of the differential co-expressed network in mouse

To validate the differential co-expression network identified in human SAT, we used adipose tissue gene expression data that were available from a mouse F2 intercross, although this was from white adipose tissue rather than pure subcutaneous adipose tissue [43]. The first and third quartiles of mouse weight were used to select the most obese and most lean mice (n=144). Orthologous genes were identified using Ensembl Biomart (build 37) [44]. For comparison, the differential co-expression analysis in humans was re-evaluated using the subset of genes also present in the mouse dataset. To assess the empirical significance of the difference observed between relevance networks, we applied the differential co-expression analysis approach at 10% FDR level. Assuming the independence of the two samples, in each permuted data set the pooled sample was randomly split preserving the original sample size of the two groups.

Statistical assessment was carried out to determine whether any gene showed a number of connections in both the human and mouse differential co-expression networks higher than expected under the assumption of independence. Assuming that the number of connections in each network follows a Poisson distribution, we simulated 1 000 000 times a sample of n paired observations from two independent Poisson, with n equal to the number of genes used to build the two networks. In each simulation we calculated the proportion of connections for the same gene in both networks and we recorded the highest joint proportion which, under the null hypothesis, corresponds to the product of the two marginal distributions. Finally, the empirical distribution of the highest joint proportion was used to evaluate the empirical P value for each pair of significant genes identified in both the human and mouse difference relevance networks.

Correlation of NEGR1 Gene expression in human SAT and hypothalamus

In order to investigate the possibility of correlation between expression of NEGR1 in adipose tissue and in the hypothalamus, a publicly available dataset was used (NCBI GEO accession number GSE3526) [45]. This study analysed gene expression in different normal tissues from ten healthy donors using the Affymetrix Human Genome U133 Plus 2.0 Array. Genome-wide expression levels in hypothalamus were available for eight subjects. For three subjects, expression levels were also available for adipose tissue. We assessed *NEGR1* correlation in expression levels between the two tissues, using the genome-wide data to generate a null distribution of no association. An empirical P value was derived using one million permutations.

Results

Differentially expressed transcripts

We determined which transcripts were differentially expressed between obese and lean subjects. The results are reported in Table 1. Obesity showed a global effect on genome-wide gene expression. A majority (55%) of the differentially expressed transcripts were up-regulated in lean subjects. DAVID/KEGG analysis of the differentially expressed transcripts did not identify significant enrichment for any obvious obesity-related pathway.

Detection of *cis*-eQTLs

Given the inter-SNP map distances, we defined a *cis*-eQTL signal for each transcript as the maximum LOD-score obtained within 2.5cM 5′ or 3′ of each transcript position in the genome. There were 26 421 transcripts with a SNP marker within 5cM. Median (1st - 3rd

quartile) heritability was 0.19 (0.05 – 0.34). The maximum LOD-score was detected at a median (1st - 3rd quartile) distance from the centre of the transcript of 1.5 cM (0.8 - 2.1cM). We identified 1 063 (4%) eQTLs at 10% FDR level. The twenty *cis*-eQTLs with the highest LOD-scores are shown in Table 2. As expected, *cis*-eQTLs were detected for those expression traits with a heritability score of zero or close to zero but traits with higher heritability also had higher LOD-scores.

Detection of *trans*-eQTLs

For each transcript we recorded the maximum peak LOD-score located on a chromosome different to the chromosome where the transcript was located. Using simulations (see Methods) with a 10% FDR, we identified 50 significant *trans*-eQTLs distributed across 12 chromosomes (see Table 3). Although most *trans*-eQTLs were not significant after multiple testing correction, we noted that *trans*-linkage signals for many transcripts were concentrated in the 1p13.3-q23.3 region. The empirical probability of observing coincident linkage was tested by simulating under the null hypothesis of no linkage. In the simulations, when a false positive was detected for a transcript, a number of correlated transcripts also showed a linkage peak in the same region, as expected. Using the empirical probability of coincident linkages through the genome, we determined a significant clustering of 374 transcripts in the 1p13.3-q23.3 region at 10% FDR.

Biological pathways involved in obesity

Given the set of 1 063 *cis*-eQTLs, pathway enrichment analysis using DAVID [46, 47] identified the KEGG insulin signalling pathway as the most significantly enriched (P=1.6×10^{-2}). The proportion of differentially expressed genes in this pathway did not differ from that observed in the whole dataset. No significant enrichment was observed for the small number of *trans*-regulated genes identified in this study. For the hotspot of *trans*-regulators on chromosome 1p13.3-q23.3 the proportion of obesity-related transcripts was again not different from what would be expected at random. Significant enrichment was observed for genes in the apoptosis pathway (P=5.5×10^{-3}) but no obvious obesity candidates were present in this very large region.

To identify obesity-related networks that include transcripts under genetic control, we focused on 425 transcripts that for an FDR of 10% were both differentially expressed between lean and obese subjects and under *cis*-acting regulation. Using an EASE [48] score threshold of 0.1 in DAVID to rank categories of genes, only the Cell Adhesion Molecule (CAMs) KEGG functional grouping was highlighted, which in our dataset contains 160 transcripts (corresponding to 76 genes), eight of them (corresponding to seven genes) under *cis*-regulation. Relevance gene networks were constructed separately in obese and lean subjects using these 160 transcripts and the empirical significance of the observed differences in co-expression among pairs of transcripts in the two networks evaluated by permutations (see Methods).

The lean and obese relevance networks and their contrast are shown in Figure 1. Table 4 lists the CAMs genes and their number of connections in the contrasted network, i.e. the number of significantly different correlations (FDR < 10%) of each gene with the remaining genes between the two groups. The neuronal growth regulator 1 (*NEGR1*) gene was the most connected gene with nine edges, while four significant connections were observed for *HLA-DQB2*, three for *ALCAM*, and two for *HLA-DQA2*, *ITGAM*, and *CD86*. We tested whether the number of connections observed for these genes was larger than that expected by chance. Using as the null distribution a Poisson random variable with mean equal to the average connectivity in the network, the 9 connections observed for *NEGR1* were

considered as a rare event (P=4.4×10^{-10}). Having four, three, and two connections in this dataset corresponds to P values of 2.7×10^{-4}, 0.002, and 0.02, respectively.

We also evaluated the empirical significance of the connectivity observed for these genes by contrasting relevance networks (between obese and lean subjects) randomly generated by using the same number of transcripts and recording the gene with highest connectivity in each simulated dataset. Out of 1 000 replicates, sporadic differences were observed between the obese and lean correlation matrices, as expected, but none of them showed a similar number of differences with respect to the original dataset. In no cases did a sample size of 160 transcripts contain a gene with nine edges. Marginal significance was observed for *HLA-DQB2* (P=0.028).

Validation of the CAMs network in mouse

From the whole set of 76 genes belonging to the human CAMs pathway, 57 orthologous genes were present in a mouse dataset [43], corresponding to 66 mouse transcripts. To assess the importance of the *NEGR1* gene in both humans and mice, we first restricted the set of CAMs genes in the human data to those which were also present in the mouse, resulting in a set of 115 human transcripts. Contrast of the co-expression networks were carried out in human and mouse and significant results filtered using a 10% FDR level. Table 5 shows the list of significantly connected genes from the mouse analysis, highlighting that *NEGR1* is highly connected in the contrasted mouse network as well. The mouse differential relevance network contained an overall larger number of connections, probably reflecting higher intra-group homogeneity and reduced environmental noise in this dataset. We ordered each gene with respect to the observed joint connectivity in both networks. The empirical significance of its rank was assessed through simulations under the null hypothesis of networks' independence (see Methods). Only *NEGR1* showed a significant departure from this assumption (P=2.1×10^{-5}) indicating that this gene is integral to both the human and mouse networks.

Expression of the *NEGR1* transcript in SAT

The NEGR1 gene, central to the contrasted co-expression network, is expressed at high levels in brain [7]. Using quantitative real-time PCR we demonstrated that *NEGR1* is also expressed in SAT (as well as heart and skeletal muscle) using a commercially-available tissue panel and two independent unrelated human SAT samples (Figure 2).

Correlation of NEGR1 expression in human SAT and hypothalamus

NEGR1 expression levels were significantly correlated between adipose and hypothalamus tissues (r = 0.99; P value = 0.020). This places *NEGR1* in the top 3% of the most correlated transcripts genome-wide. We also assessed the empirical significance of our finding using genome-wide expression data in the two tissues to generate a null distribution of no association (empirical P value = 0.022).

Discussion

An important goal of systems biology is the identification of biological pathways and genetic networks underlying complex human diseases. We studied genome-wide gene expression in SAT and its genetically determined variation in families ascertained through sib-pairs discordant for obesity. The expression of about 30% of all genes was significantly altered in the obese state, confirming a broad effect of obesity on SAT gene expression [21].

Linkage analysis identified a large number of significant eQTL, most of them localized in *cis*, and a lesser number of *trans*-acting signals perhaps due to reduced power of detection.

Gene Ontology and pathway analyses of the *cis*-regulated genes demonstrated that they were enriched for genes involved in the insulin signalling pathway. The identification of genetic regulation of the insulin pathway is intriguing as it may indicate a role for SAT in glucose homeostasis and identify its contribution to the development of polygenic type 2 diabetes [49]. Clear identification of genetic regulation of this pathway in SAT suggests that exploration of the regulated genes may give valuable insights into the fact that only a minority of obese subjects develop T2D, and those that do, typically have insulin resistance, metabolic syndrome and insulin secretion defects [50]. No significant biological clustering was observed for the small number of *trans*-regulated genes identified in this study. A group of 374 transcripts suggested the presence of a significant hotspot for *trans*-regulation on chromosome 1p13.3-q23.3 and it may be of note that this overlaps the well-replicated T2D linkage locus of 1q21-q25 [51]. Significant enrichment was observed for genes in the apoptosis pathway but no obvious candidates could be identified in this very large region.

While differential expression analysis can identify those genes and pathways with a causal or reactive role in obesity, genetic analysis can highlight which of them are under genetic control and therefore likely to be "functionally" transcribed in SAT cells [52]. Therefore, whereas differential expression may be a result of the "obesogenic environment", those biological pathways enriched for differentially expressed and genetically controlled genes are more likely to have a causative role in the development of obesity. The subset of *cis*-regulated transcripts which were also differentially expressed between lean and obese subjects suggested a possible role for genes belonging to the CAMs functional grouping. Contrasting CAMs co-expression networks between lean and obese subjects identified a subset of genes whose pattern of co-expression was significantly associated with the obese state. We found *NEGR1* as the central highly-connected gene of this subset and replicated this observation using a mouse expression dataset, thus validating its central role for this pathway. In the context of disease, the topology of gene networks is often determined by key genes showing a high degree of connectivity. Indeed, highly connected genes are likely to encode essential genes [53] which are often evolutionarily conserved [54]. Genes showing an intermediate number of connections have been shown to be more likely to harbour inherited mutations for common diseases [55]. Whereas *NEGR1* was the most connected gene in the contrasted network, it showed intermediate connectivity within each group-specific co-expression network, thus supporting its possible role as a disease gene. The GiANT consortium meta-analysis of obesity GWA studies reported that genetic markers near the *NEGR1* gene are associated with obesity [7].

The NEGR1 protein is a member of the immunoglobulin superfamily, is highly expressed in the hypothalamus [56] where it appears to modulate synapse number in neurons [57] and this makes it a good functional candidate for obesity [58], especially when considering obesity as a disorder having a neurobehavioral origin [59]. Our findings demonstrate that *NEGR1* is expressed in human SAT where it appears to be central to the network of the most differentially-expressed set of functionally-related genes between lean and obese subjects. Using publicly-available data [60] we observed high correlation in the expression levels of *NEGR1* between human subcutaneous adipose and hypothalamus tissues. These results suggest a similar expression pattern for *NEGR1* across tissues. Thus, transcriptional regulation of *NEGR1* might not be restricted to neural development and might involve additional mechanisms shared by other tissues.

In addition to *NEGR1*, other genes in the CAMs network have been previously shown to be over-expressed in SAT. In a study of BMI-discordant identical twins [61] the up-regulation of inflammatory and cytoskeleton pathways and down-regulation of energy metabolism and cell differentiation pathways was clearly demonstrated. Specifically, an over-expression of MHC Class II transcripts in obese subjects was reported and these are present in our CAMs

relevance network. This further supports the utility of our approach and suggests that other genes within the identified obesity subset of CAMs genes might be good candidates for further investigation.

In summary, we have identified a subset of genes that are both differentially-expressed between lean and obese subjects and are under *cis*-regulation, and so are very good candidates to investigate further for the presence of gene variants regulating their expression and thus contributing to obesity. We have applied a novel differential co-expression analysis strategy to identify *NEGR1* as a gene central to the CAMs network in the obese state and confirmed this finding in a different species.

Acknowledgments

The authors wish to acknowledge the participation of all the families and clinical staff involved in the SOS SibPair study.

The authors thank Prof. Eric Schadt for advice and the provision of the mouse dataset and the staff of the Imperial College High-Performance Computing Service for their advice and support.

This study was funded by grant no. 079534/z/06/z from the Wellcome Trust, the Swedish Research Council (K2010-55X-11285-13), the Swedish foundation for Strategic Research to Sahlgrenska Center for Cardiovascular and Metabolic Research, the Swedish Diabetes foundation and the Swedish federal government under the LUA/ALF agreement. Sylvia Richardson acknowledges support from the MRC grant G0600609.

REFERENCES

1. Kelly T, Yang W, Chen CS, Reynolds K, He J. Global burden of obesity in 2005 and projections to 2030. Int J Obes (Lond). 2008; 32(9):1431–7. [PubMed: 18607383]
2. Haslam DW, James WP. Obesity. Lancet. 2005; 366(9492):1197–209. [PubMed: 16198769]
3. Walley AJ, Asher JE, Froguel P. The genetic contribution to non-syndromic human obesity. Nat Rev Genet. 2009; 10(7):431–42. [PubMed: 19506576]
4. Saunders CL, Chiodini BD, Sham P, Lewis CM, Abkevich V, Adeyemo AA, et al. Meta-analysis of genome-wide linkage studies in BMI and obesity. Obesity (Silver Spring). 2007; 15(9):2263–75. [PubMed: 17890495]
5. Hinney A, Nguyen TT, Scherag A, Friedel S, Bronner G, Muller TD, et al. Genome wide association (GWA) study for early onset extreme obesity supports the role of fat mass and obesity associated gene (FTO) variants. PLoS One. 2007; 2(12):e1361. [PubMed: 18159244]
6. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet. 2009; 41(1):18–24. [PubMed: 19079260]
7. Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nat Genet. 2009; 41(1): 25–34. [PubMed: 19079261]
8. Meyre D, Delplanque J, Chevre JC, Lecoeur C, Lobbens S, Gallina S, et al. Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. Nat Genet. 2009; 41(2):157–9. [PubMed: 19151714]
9. Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J, et al. A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. Nature. 2010; 463(7281):671–5. [PubMed: 20130649]
10. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, et al. Genetics of gene expression surveyed in maize, mouse and man. Nature. 2003; 422(6929):297–302. [PubMed: 12646919]
11. Idaghdour Y, Storey JD, Jadallah SJ, Gibson G. A genome-wide gene expression signature of environmental geography in leukocytes of Moroccan Amazighs. PLoS Genet. 2008; 4(4):e1000052. [PubMed: 18404217]
12. Schadt EE. Molecular networks as sensors and drivers of common human diseases. Nature. 2009; 461(7261):218–23. [PubMed: 19741703]

13. Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. Mapping complex disease traits with global gene expression. Nat Rev Genet. 2009; 10(3):184–94. [PubMed: 19223927]
14. Petretto E, Mangion J, Dickens NJ, Cook SA, Kumaran MK, Lu H, et al. Heritability and tissue specificity of expression quantitative trait loci. PLoS Genet. 2006; 2(10):e172. [PubMed: 17054398]
15. Gerrits A, Li Y, Tesson BM, Bystrykh LV, Weersing E, Ausema A, et al. Expression quantitative trait loci are highly sensitive to cellular differentiation state. PLoS Genet. 2009; 5(10):e1000692. [PubMed: 19834560]
16. Wellen KE, Hotamisligil GS. Obesity-induced inflammatory changes in adipose tissue. J Clin Invest. 2003; 112(12):1785–8. [PubMed: 14679172]
17. Vazquez-Vela ME, Torres N, Tovar AR. White adipose tissue as endocrine organ and its role in obesity. Arch Med Res. 2008; 39(8):715–28. [PubMed: 18996284]
18. van Beek EA, Bakker AH, Kruyt PM, Hofker MH, Saris WH, Keijer J. Intra- and interindividual variation in gene expression in human adipose tissue. Pflugers Arch. 2007; 453(6):851–61. [PubMed: 17061120]
19. Jiao H, Kaaman M, Dungner E, Kere J, Arner P, Dahlman I. Association analysis of positional obesity candidate genes based on integrated data from transcriptomics and linkage analysis. Int J Obes (Lond). 2008; 32(5):816–25. [PubMed: 18180783]
20. Chen Y, Zhu J, Lum PY, Yang X, Pinto S, MacNeil DJ, et al. Variations in DNA elucidate molecular networks that cause disease. Nature. 2008; 452(7186):429–35. [PubMed: 18344982]
21. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, et al. Genetics of gene expression and its effect on disease. Nature. 2008; 452(7186):423–8. [PubMed: 18344981]
22. Carlsson LM, Jacobson P, Walley A, Froguel P, Sjostrom L, Svensson PA, et al. ALK7 expression is specific for adipose tissue, reduced in obesity and correlates to factors implicated in metabolic disease. Biochem Biophys Res Commun. 2009; 382(2):309–14. [PubMed: 19275893]
23. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet. 1998; 63(1):259–66. [PubMed: 9634505]
24. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet. 2002; 30(1):97–101. [PubMed: 11731797]
25. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003; 4(2):249–64. [PubMed: 12925520]
26. Poulain-Godefroy O, Lecoeur C, Pattou F, Fruhbeck G, Froguel P. Inflammation is associated with a decrease of lipogenic factors in omental fat in women. Am J Physiol Regul Integr Comp Physiol. 2008; 295(1):R1–7. [PubMed: 18448614]
27. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet. 1998; 62(5):1198–211. [PubMed: 9545414]
28. Box GEP, Cox DR. An analysis of transformations. J R Stat Soc B. 1964; 26:211–252.
29. Abecasis GR, Wigginton JE. Handling marker-marker linkage disequilibrium: pedigree analysis with clustered markers. Am J Hum Genet. 2005; 77(5):754–67. [PubMed: 16252236]
30. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B. 1995; 57:289–300.
31. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004; 3 Article3.
32. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003; 100(16):9440–5. [PubMed: 12883005]
33. Kleinjan DA, van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet. 2005; 76(1):8–32. [PubMed: 15549674]
34. Li KC. Genome-wide coexpression dynamics: theory and application. Proc Natl Acad Sci U S A. 2002; 99(26):16875–80. [PubMed: 12486219]
35. Choi JK, Yu U, Yoo OJ, Kim S. Differential coexpression analysis using microarray data and its application to human cancer. Bioinformatics. 2005; 21(24):4348–55. [PubMed: 16234317]

36. Zhu D, Hero AO, Cheng H, Khanna R, Swaroop A. Network constrained clustering for gene microarray data. Bioinformatics. 2005; 21(21):4014–20. [PubMed: 16141248]
37. Pesarin, F. Multivariate Permutation Tests : With Applications in Biostatistics. Wiley; 2001.
38. Fang G, Kuang R, Pandey G, Steinbach M, Myers CL, Kumar V. Subspace differential coexpression analysis: problem definition and a general approach. Pac Symp Biocomput. 2010:145–56. [PubMed: 19908367]
39. Fuller TF, Ghazalpour A, Aten JE, Drake TA, Lusis AJ, Horvath S. Weighted gene coexpression network analysis strategies applied to mouse weight. Mamm Genome. 2007; 18(6-7):463–72. [PubMed: 17668265]
40. Xu M, Kao MC, Nunez-Iglesias J, Nevins JR, West M, Zhou XJ. An integrative approach to characterize disease-specific pathways and their coordination: a case study in cancer. BMC Genomics. 2008; 9(Suppl 1):S12. [PubMed: 18366601]
41. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. Proc Natl Acad Sci U S A. 2006; 103(47):17973–8. [PubMed: 17101986]
42. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004; 5(2):101–13. [PubMed: 14735121]
43. Wang S, Yehya N, Schadt EE, Wang H, Drake TA, Lusis AJ. Genetic and genomic analysis of a fat mass trait with complex inheritance reveals marked sex specificity. PLoS Genet. 2006; 2(2):e15. [PubMed: 16462940]
44. Kasprzyk A, Keefe D, Smedley D, London D, Spooner W, Melsopp C, et al. EnsMart: a generic system for fast and flexible access to biological data. Genome Res. 2004; 14(1):160–9. [PubMed: 14707178]
45. Roth RB, Hevezi P, Lee J, Willhite D, Lechner SM, Foster AC, et al. Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. Neurogenetics. 2006; 7(2): 67–80. [PubMed: 16572319]
46. Dennis G Jr. Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 2003; 4(5):P3. [PubMed: 12734009]
47. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protocols. 2009; 4(1):44–57.
48. Hosack DA, Dennis G Jr. Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE. Genome Biol. 2003; 4(10):R70. [PubMed: 14519205]
49. Frojdo S, Vidal H, Pirola L. Alterations of insulin signaling in type 2 diabetes: a review of the current evidence from humans. Biochim Biophys Acta. 2009; 1792(2):83–92. [PubMed: 19041393]
50. Iozzo P. Viewpoints on the way to the consensus session: where does insulin resistance start? The adipose tissue. Diabetes Care. 2009; 32(Suppl 2):S168–73. [PubMed: 19875546]
51. Prokopenko I, Zeggini E, Hanson RL, Mitchell BD, Rayner NW, Akan P, et al. Linkage disequilibrium mapping of the replicated type 2 diabetes linkage signal on chromosome 1q. Diabetes. 2009; 58(7):1704–9. [PubMed: 19389826]
52. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science. 2009; 325(5945):1246–50. [PubMed: 19644074]
53. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. Proc Natl Acad Sci U S A. 2007; 104(21):8685–90. [PubMed: 17502601]
54. Bergmann S, Ihmels J, Barkai N. Similarities and differences in genome-wide expression data of six organisms. PLoS Biol. 2004; 2(1):E9. [PubMed: 14737187]
55. Feldman I, Rzhetsky A, Vitkup D. Network properties of genes harboring inherited disease mutations. Proc Natl Acad Sci U S A. 2008; 105(11):4323–8. [PubMed: 18326631]
56. Miyata S, Funatsu N, Matsunaga W, Kiyohara T, Sokawa Y, Maekawa S. Expression of the IgLON cell adhesion molecules Kilon and OBCAM in hypothalamic magnocellular neurons. J Comp Neurol. 2000; 424(1):74–85. [PubMed: 10888740]

57. Hashimoto T, Yamada M, Maekawa S, Nakashima T, Miyata S. IgLON cell adhesion molecule Kilon is a crucial modulator for synapse number in hippocampal neurons. Brain Res. 2008; 1224:1–11. [PubMed: 18602091]

58. Bauer F, Elbers CC, Adan RA, Loos RJ, Onland-Moret NC, Grobbee DE, et al. Obesity genes identified in genome-wide association studies are associated with adiposity measures and potentially with nutrient-specific food preference. Am J Clin Nutr. 2009; 90(4):951–9. [PubMed: 19692490]

59. O'Rahilly S, Farooqi IS. Human obesity: a heritable neurobehavioral disorder that is highly sensitive to environmental conditions. Diabetes. 2008; 57(11):2905–10. [PubMed: 18971438]

60. Yang X, Deignan JL, Qi H, Zhu J, Qian S, Zhong J, et al. Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks. Nat Genet. 2009; 41(4):415–23. [PubMed: 19270708]

61. Pietilainen KH, Naukkarinen J, Rissanen A, Saharinen J, Ellonen P, Keranen H, et al. Global transcript profiles of fat in monozygotic twins discordant for BMI: pathways behind acquired obesity. PLoS Med. 2008; 5(3):e51. [PubMed: 18336063]

Europe PMC Funders Author Manuscripts

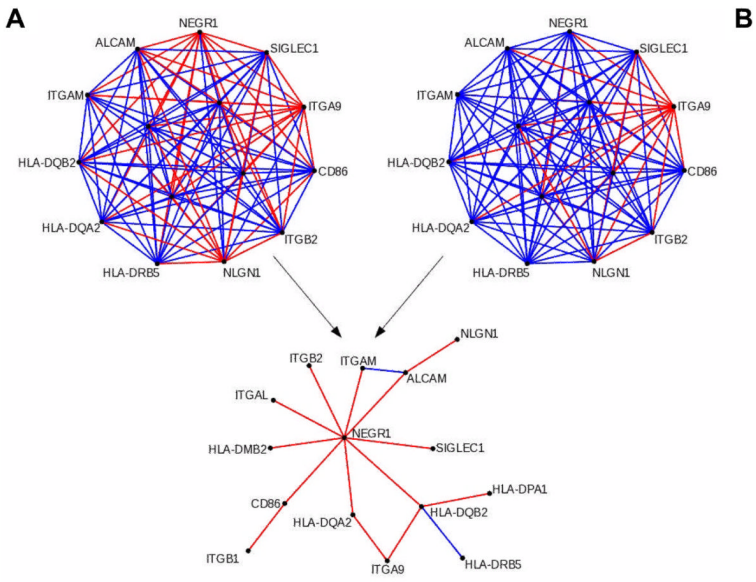

Figure 1. Differential co-expression analysis of CAM gene expression in human subcutaneous adipose tissue

Differentially co-expressed network of the CAMs functional grouping resulting from the contrast between obese (A) and lean (B) networks at FDR 10%. Red and blue edges represent negative and positive correlations respectively. For simplicity, gene names are only shown for the external nodes in (A) and (B).

Europe PMC Funders Author Manuscripts

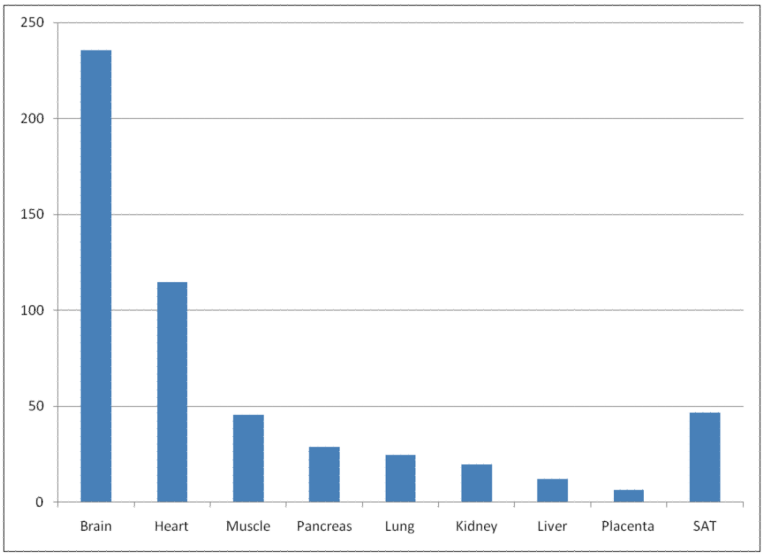

Figure 2. *NEGR1* expression levels in human tissues

Relative expression of the *NEGR1* transcript in human subcutaneous adipose tissue (SAT) compared to expression in other human tissues from a commercially-available multiple tissue panel.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Numbers (and percentage) of differentially expressed transcripts between lean and obese subjects identified using the Limma package at different FDR levels using the linear regression option. Number (and percentage) of upregulated transcripts in obese subjects is also provided.

FDR level	Transcripts differentially expressed (%)	Upregulated in obese (%)
5%	12 621 (23)	6 179 (49)
10%	16 454 (30)	7 478 (45)
20%	23 251 (43)	9 679 (42)

Europe PMC Funders Author Manuscripts

Int J Obes (Lond). Author manuscript; available in PMC 2012 July 01.

Table 2

Top twenty human *cis*-eQTLs across the whole genome in descending order of LOD-score. Chromosomal position is based upon NCBI build 37 of the human genome.

Affymetrix ID	Gene Symbol	Gene Name	Chr	LOD	Position (Mb)	Distance (Mb)
219629_at	*FAM118A*	family with sequence similarity 118, member A	22	10.9	43.59	0.51
228425_at	*LOC654433*	hypothetical LOC654433	2	10.4	113.70	0.02
203815_at	*GSTT1*	glutathione S-transferase theta 1	22	10.3	23.55	0.84
222424_s_at	*NUCKS1*	nuclear casein kinase and cyclin-dependent kinase substrate 1	1	9.0	203.13	0.84
219269_at	*HMBOX1*	homeobox containing 1	8	9.0	28.82	0.06
226707_at	*NAPRT1*	nicotinate phosphoribosyltransferase domain containing 1	8	8.1	144.83	0.10
214012_at	*ERAP1*	endoplasmic reticulum aminopeptidase 1	5	7.9	94.88	1.26
209255_at	*KLHDC10*	kelch domain containing 10	7	7.7	128.49	1.04
203337_x_at	*ITGB1BP1*	integrin beta 1 binding protein 1	2	7.5	8.10	1.37
224097_s_at	*F11R*	F11 receptor	1	7.4	158.52	0.74
203814_s_at	*NQO2*	NAD(P)H dehydrogenase, quinone 2	6	7.3	2.84	0.12
224828_at	*CPEB4*	cytoplasmic polyadenylation element binding protein 4	5	7.3	173.27	0.01
226736_at	*CHURC1*	churchill domain containing 1	14	7.3	66.22	1.76
225278_at	*PRKAB2*	protein kinase, AMP-activated, beta 2 non-catalytic subunit	1	7.1	143.21	1.89
224904_at	*PDPR*	pyruvate dehydrogenase phosphatase regulatory subunit	16	7.1	68.01	0.71
203096_s_at	*RAPGEF2*	Rap guanine nucleotide exchange factor (GEF) 2	4	6.5	160.55	0.09
219027_s_at	*MYO9A*	myosin IXA	15	6.5	72.03	1.98
227678_at	*XRCC6BP1*	XRCC6 binding protein 1	12	6.5	55.62	1.01
229238_at	*C17orf97*	chromosome 17 open reading frame 97	17	6.5	1.57	1.30
210947_s_at	*MSH3*	mutS homolog 3 (E. coli)	5	6.5	78.71	1.39

Europe PMC Funders Author Manuscripts

Table 3

List of human *trans*-eQTLs across the whole genome in chromosomal position order and descending order of LOD-score where *trans*-eQTLs have the same position. Chromosomal position is based upon NCBI build 37 of the human genome.

Affymetrix ID	Gene Symbol	Gene Name	Chr	LOD	Position (Mb)
204440_at	*CD83*	CD83 molecule	1	13.4	53.22
227610_at	*TSPAN11*	Tetraspanin 11	1	13.1	53.22
204973_at	*GJB1*	Gap junction protein, beta 1	1	8.6	53.22
205203_at	*PLD1*	Phospholipase D1, phosphatidylcholine-specific	1	11.9	109.84
208631_s_at	*HADHA*	Hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), alpha subunit	1	13.4	113.22
217169_at	*IGHA1*	Immunoglobulin heavy constant alpha 1	1	8.8	115.34
225279_s_at	*C3ORF17*	Chromosome 3 open reading frame 17	1	9.2	120.23
200874_s_at	*NOL5A*	Nucleolar protein 5A	1	14.8	143.21
222735_at	*TMEM38B*	Transmembrane protein 38B	1	8.9	148.37
224755_at	*TM9SF3*	Transmembrane 9 superfamily member 3	1	11.7	157.05
200894_s_at	*FKBP4*	FK506 binding protein 4	1	15.1	169.80
222425_s_at	*POLDIP2*	Polymerase (DNA-directed), delta interacting protein 2	1	8.9	203.13
223193_x_at	*C3ORF28*	Chromosome 3 open reading frame 28	1	9.9	212.88
217959_s_at	*TRAPPC4*	Trafficking protein particle complex 4	1	8.3	212.88
228451_at	*TSSK3*	Testis-specific serine kinase 3	1	9.1	227.81
202000_at	*NDUFA6*	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 6	1	9.6	228.89
227607_at	*STAMBPL1*	STAM binding protein-like 1	2	10.1	38.12
228426_at	*CLEC2D*	C-type lectin domain family 2, member D	2	10.4	113.70

Europe PMC Funders Author Manuscripts

Table 4

Genes showing significantly different co-expression between lean and obese human SAT CAMs networks at FDR 10% in descending order of their number of connections

Gene Symbol	Gene Name	Connectio ns
NEGR1	neuronal growth regulator 1	9
HLA-DQB2	major histocompatibility complex, class II, DQ beta 2	4
ALCAM	activated leukocyte cell adhesion molecule	3
CD86	CD86 antigen	2
HLA-DQA2	major histocompatibility complex, class II, DQ alpha 2	2
ITGA9	integrin, alpha 9	2
ITGAM	integrin, alpha M (complement component 3 receptor 3 subunit)	2
SELP	selectin P (granule membrane protein 140kDa, antigen CD62)	2
CADM1	cell adhesion molecule 1	1
CD99	CD99 antigen	1
GLG1	golgi apparatus protein 1	1
HLA-DMB	major histocompatibility complex, class II, DM beta	1
HLA-DPA1	major histocompatibility complex, class II, DP alpha 1	1
HLA-DRB5	major histocompatibility complex, class II, DR beta 5	1
ITGA8	integrin, alpha 8	1
ITGAL	integrin, alpha L	1
ITGB1	integrin, beta 1	1
ITGB2	integrin, beta 2	1
NLGN1	neuroligin 1	1
NRCAM	neuronal cell adhesion molecule	1
NRXN1	neurexin 1	1
PTPRM	protein tyrosine phosphatase, receptor type, M	1
PVRL3	poliovirus receptor-related 3	1
SIGLEC1	sialic acid binding Ig-like lectin 1, sialoadhesin	1

Europe PMC Funders Author Manuscripts

Table 5

Genes showing significantly different co-expression by contrasting the mouse SAT CAMs networks between mice in the first and third quartile of the weight distribution at 10% FDR level in descending order of their number of connections.

Gene symbol	Gene name	Connectio ns
SDC2	syndecan 2	25
CLDN11	claudin 11	23
CDH2	cadherin 2	22
PTPRF	protein tyrosine phosphatase, receptor type, F	21
NEGR1	neuronal growth regulator 1	20
NRXN1	neurexin I	19
F11R	F11 receptor	18
NLGN1	Neuroligin 1	17
HLA-DRB5	major histocompatibility complex, class II, DR beta 5	16
ITGB1	integrin beta 1	16
JAM2	junction adhesion molecule 2	16
PTPRM	protein tyrosine phosphatase, receptor type, M	14
SIGLEC1	sialic acid binding Ig-like lectin 1, sialoadhesin	13
VCAN	versican	13
ALCAM	activated leukocyte cell adhesion molecule	12
ITGB8	integrin beta 8	12
NRXN3	neurexin 3	12
CLDN15	claudin 15	11
ICAM1	intercellular adhesion molecule	10
JAM3	junction adhesion molecule 3	10
SPN	sialophorin	10
CADM1	cell adhesion molecule 1	9
PTPRC	protein tyrosine phosphatase, receptor type, C	9
CLDN5	claudin 5	8
ITGA8	integrin alpha 8	8
ITGAL	integrin alpha L	8
ITGAM	integrin alpha M	8
ITGAV	integrin alpha V	8
SELPLG	selectin, platelet (p-selectin) ligand	8
CD274	CD274 antigen	7
CD86	CD86 antigen	7
CLDN19	claudin 19	7
HLA-DQB2	major histocompatibility complex, class II, DQ beta 2	7
ITGA9	integrin alpha 9	7
MPZL1	myelin protein zero-like 1	7
NCAM1	neural cell adhesion molecule 1	7

Europe PMC Funders Author Manuscripts

Gene symbol	Gene name	Connectio ns
CD2	CD2 antigen	6
HLA-DOA	major histocompatibility complex, class II, DO alpha	6
SELP	selectin, platelet	6
VCAM1	vascular cell adhesion molecule 1	6
CD4	CD4 antigen	5
CLDN9	claudin 9	5
NEO1	neogenin	5
PVRL3	poliovirus receptor-related 3	5
GLG1	golgi apparatus protein 1	4
PVRL2	poliovirus receptor-related 2	4
NCAM2	neural cell adhesion molecule 2	3
NRCAM	neuron-glia-CAM-related cell adhesion molecule	3
CD34	CD34 antigen	2
ITGB2	integrin beta 2	2
NLGN3	neuroligin 3	2
CD28	CD28 antigen	1
CD8B	CD8b antigen	1
PDCD1	programmed cell death 1	1